SPATIAL REGRESSION OF CROP PARAMETERS WITH AIRBORNE SPECTRAL IMAGERY*

D.J.Timlin
USDA-ARS Alternate Crops and Systems Laboratory
Beltsville, MD USA 20705
C.L. Walthall, Ya. Pachepsky, W.P. Dulaney, C.S.T. Daughtry
USDA-ARS Hydrology and Remote Sensing Laboratory
Beltsville, MS USA 20705

ABSTRACT

The spatial information in remotely sensed images is often not fully exploited to improve calibrations of image tone with plant factors. A number of statistical tools that include ordinary least squares regression (OLS) are used to develop equations that relate image tone to plant leaf area or biomass. Methods such as OLS ignore the spatial relationships in the data which may result in erroneous conclusions. Further, these methods generally provide a global fit to the data. Methods such as Spatial Autoregression (SAR) that correct for spatial autocorrelation in errors can be more effective in accounting for local information when developing calibration equations. We used data from an airborne image with a 1m X 1m resolution on a 4.5 ha field taken in July, 1998. Leaf area index (LAI) was measured in a corn crop at 74 locations on an irregular grid concurrent with image aquisition. An autoregressive model that uses a 2 dimensional spatial weighting matrix was used to relate the Modified Soil and Vegetation Index (MSAVI) calculated from the image pixels to LAI. MSAVI was averaged over 3, 5, 7 and 11 meter blocks. We compared the AR model to ordinary least squares (OLS) regression that does not account for spatial information. The AR model provided a better fit to the LAI data than the OLS model and increased the r^2 from 29 to 34%. The best fit was obtained using 5-m block averages.

1.0 INTRODUCTION

The use of remotely sensed data in agricultural enterprises has the potential to provide high density and timely spatial information on crop condition. Remotely sensed data can also useful information on temporal changes in crop condition.

The benefits of using remotely sensed data to monitor crop growth are more effectively realized when quantitative variables such as foliage density (estimated as Leaf Area Index, LAI) can be estimated from the image. Benedetti and Rossini (1993) used a linear model to forecast wheat yield based on a normalized vegetation density index (NDVI) integration during the grain filling period.

Bothe the remotely sensed data and measured LAI however, are often spatially correlated. The presence of autocorrelation is recognized (Atkinson and Curran, 1997), but efforts to utilize this information are still not widespread. Autocorrelation can be used to advantage since knowledge of nearby values of LAI can help improve the relationship between LAI and a vegetation index from the image.

Relationships between image data, such as the normalized vegetation index (NDVI), and LAI are empirical and require the use of regression to determine the parameters (Benedetti and Rossini, 1993). NDVI and LAI are generally spatially correlated over space (Atkinson and Curran, 1997). This means that if a large value of NDVI or LAI is recorded in one location, then it is highly likely that a neighboring location will also have a high value. When ordinary least squares (OLS) regression is carried out, the errors will be spatially autocorrelated if there is

*Presented at the Third International Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado, 5-7 November 2001.

autocorrelation in the input data. The spatial autocorrelation can result in biased estimators from the OLS regression. This is because some of the spatial effect may be assigned to one of the predictors in OLS regression making that predictor seem more important than it really is. Long (1998) has shown how the level of significance for the F statistic can be greatly inflated when spatial autocorrelation is not taken into account in OLS regression. At the same time Pace and Gilley (1997) showed that by taking spatial autocorrelation into account, some predictors take on a higher level of significance since local response is considered rather than a global optimization. Chica-Olmo and Abarca-Hernández., (2000) used spatial autocorrelation in their data to improve a classification method. Rather than fitting a global relationship where the parameters have to account for the average of overall variability, methods that take autocorrelation into account can better use local information on variability.

Autoregressive methods are commonly used to model time series data. Often, a value of the dependent variable at a previous time step (lag) can be used along with an independent variable to predict the value at the current time. Spatial autoregressive models have been used for data that are taken in series over space (Long, 1998). The spatial autoregression methods do not always require that the dependent and independent variables be available on regular grids having the same scales.

The objective of this study was to examine the use of spatial autoregression (SAR) to determine relationships between AISA spectral imagery and foliage density (leaf area index- LAI).

2.0 MATERIALS AND METHODS

2.1 SITE DESCRIPTION

The experiment site was a 4.5 ha corn field located on the USDA Agricultural Research Service Beltsville Agricultural Research Center in Beltsville, Maryland. The data set described here was collected on July 30, 1998. Prior to this study, the field had been in alfalfa (<u>Medicago sativa</u> L.) for 8 years (1989-1997). In the spring of 1997 the field was sprayed with herbicide to kill the alfalfa, and corn (<u>Zea Mays</u> L.) was planted into the residue using a six-row no-till planter. The row spacing was 76.2 cm and plant population about 50,000 plants ha^{-1}. Nitrogen was applied (30 days) after planting at the rate of 100 kg $ha^{-1.}$ No irrigation water was applied. The corn was planted north to south in 0.74 m wide rows. The field was divided into 64 grid cells approximately 30 m wide (E-W direction) by 18 m long (N-S direction). Four 30-m transects (six corn rows wide) were also laid out in areas with contrasting soil texture. Alleyways approximately 1 m wide were cut in an east-west position every 30 m for ease of access to the interior field locations. Differences in the vigor of the corn due to variation in nitrogen were not evident because water was severely limiting as the result of an extended drought that began soon after nitrogen application. At the time of the data collection there were no known major pest infestations.

The soil cover of the site is defined as the Cedartown-Galestown-Matawan soil association (Kirby et al., 1967). Cedartown and Galestown are sandy, siliceous, mesic Psammentic Hapludults, whereas Matawan is fine loamy, semiactive, mesic Aquic Hapludult. The slopes range from 0 to 5 %, and the topsoil texture is mostly loamy sand. The study site has a gentle slope running from the northwest of the site to southeast with an elevation difference of approximately 3 m. Sandy loam soils predominate with clay content increasing down slope.

Mean leaf area index (LAI), mean leaf tip angle (MTA), and leaf reflectance and transmittance were measured at each of the grid locations. LAI was also measured on three transects. There were eight measurements on each transect 4 meters apart (every other plot). The LAI and MTA were collected using a LICOR LAI-2000 Plant Canopy Analyzer at four locations around the central location of each grid cell.

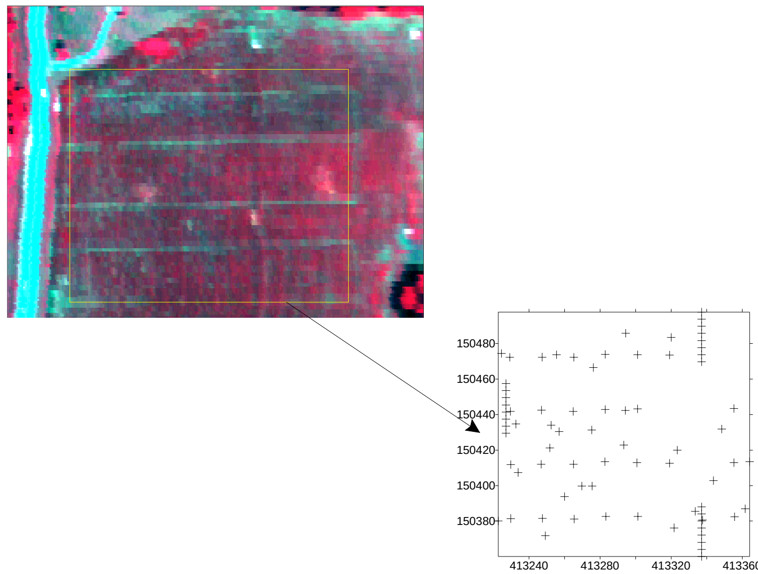

Figure A.1 AISA Image for the Field. The Boundaries of the Studied Area Are Marked. The Sampled Locations Are Indicated by Crosses.

2.2 IMAGERY

Airborne data (Figure 1) were collected using the AISA hyperspectral imaging scanner configured to record data for 26 spectral bands ranging from 434 nm to 892 nm (Braam et al., 1993). For the July 30 data acquisition, the instantaneous field of view of the scanner was 0.5 m at nadir across track and 2.5 m along track. The imagery was collected under clear sky conditions at approximately 10:30 AM EDT at an altitude of approximately 500 m above ground level. Image processing included nearest-neighbor resampling to achieve the spatial resolution and projection of the images into the Maryland State Plane Coordinate System. Spectral bands from the airborne data corresponding to those of the leaf reflectance data were subset for analysis. Both the leaf reflectance and AISA imagery bandwidths are approximately 8 to 12 nm wide. The horizontal lines in Figure 1 are 1- m wide alley ways. The Modified Soil and Vegetation Index (MSAVI) was calculatled as (Qi et al., 1994):

$$MSAVI = 0.5\ [2\ RED + 1 - \sqrt{(2\ RED + 1)^2 - 8\ (RED - NIR)}] \qquad \textbf{(1)}$$

AISA bands 15 and 20 were used for the red and NIR (near infrared) respectively.

2.3 SPATIAL AUTOREGRESSION

A brief description of spatial autoregressive models is given here. A more complete description of autoregressive models applied to spatial data can be found in Pace and Barry (1997), Long (1998) and Griffith (1990, 1988). In ordinary least squares regression, the dependent variable is a function of the independent variable as $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where **Y** and **X** are $n \times 1$ vectors, $\boldsymbol{\beta}$ is a $1 \times k$ vector of coefficients and $\boldsymbol{\epsilon}$ is error with mean of 0 and $\boldsymbol{\sigma}^2$=1. There are n observations(dependent, **Y**) and k predictors (independent, **X**). If **Y** or **X** exhibit spatial autocorrelation, then they must be corrected for values of **X** or **Y** at nearby locations. Since the data are collected on a two-dimensional grid, a value of **X** or **Y** at a lag of i-1 cannot be used to correct for autocorrelation as it can for a one dimensional series of data. Instead, a connectivity matrix **D** is used (Long. 1998, Pace and Gilley, 1997, Griffith, 1988). The autoregressive model becomes:

$$Y - \alpha DY = \beta_0 + X\beta_1 - DX\beta_2 \ + \epsilon \qquad \textbf{(2)}$$

The connectivity matrix, D, is an *n* by *n* weighting matrix with 0's on the diagonal so that only neighboring values and not the value itself are used for predictions. The value α is the autoregressive parameter and lies between 0 and 1. The error, ϵ, is distributed as $\epsilon \sim N(0,\sigma^2 I)$ where I is an identity matrix. We used an *n x n* connectivity matrix described by Pace and Gilley (1997) . The properties of this matrix are 1) the diagonals are 0, and 2) the rows sum to one. The spatial connectivity matrix can be adapted for many different kinds of weighting schemes (Pace and Gilley, 1997).

A

1	2	3
4	5	6
7	8	9

B

Cell	1	2	3	4	5	6	7	8	9
1	0	1/2	0	1/2	0	0	0	0	0
2	1/3	0	1/3	0	1/3	0	0	0	0
3	0	1/2	0	0	0	1/2	0	0	0
4	1/2	0	0	0	0	0	1/2	0	0
5	0	1/2	0	0	0	0	0	1/2	0
6	0	0	1/2	0	0	0	0	0	1/2
7	0	0	0	1	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	1	0	0	0

Figure A.2 Representation of the Spatial Weighting Matrix (*D*) for a 3 X 3 Grid. The Grid Is in a and the Weighting Matric in B. the Weights (B) Sum to 1 in the Rows.

Figure 2 shows the relationship between a 3 x 3 grid (A) with equally spaced cells and the spatial weights (B) calculated by the nearest neighbor method. For example, the first row of the adjusted Y (**αDY**), Y_{1a}, would be calculated as $\alpha 0.5Y_2 + \alpha 0.5Y_4$.

The autoregressive parameter, α is solved using maximum likelihood computations (Pace and Barry, 1997). The profile likelihood function is defined as:

$$L(\beta,\alpha,\sigma^2) = C \ + \ \ln|I - \alpha D| - \left(\frac{n}{2}\right)\ln(SSE) \qquad \textbf{(3)}$$

where C is a constant and $ln|I\text{-}\alpha D|$ is the log determinant. The SSE is the sum of squared errors:

$$SSE = (Y - \alpha DY - X\beta_{\alpha})^{t}(Y - \alpha DY - X\beta_{\alpha}) \quad \textbf{(4)}$$

where 't' denotes transpose of the matrix and β_{α} is an $n \times 1$ vector of coefficients (β_1 and β_2 in Eq[4]). Pace and Barry (1997) have developed an efficient and rapid method to evaluate the maximum of the profile likelihood function and hence solve for α and β. The log determinant, $ln|I\text{-}\alpha D|$, is calculated for a number (usually 100) values of α and then a lookup table is used to find the minimum of L in Eq. [2]. Matlab (The Mathworks, Natick, MA, USA) programs from the SpatialToolBox, v1.1[**] were used to perform the calculations.

After α was determined, spatially adjusted values of LAI were calculated by subtracting αDY from LAI. These adjusted LAI values were next used in an OLS regression with MSAVI (X) and its spatial lag (XD). The sums of squares for α were calculated by subtracting the total sums of squares for the OLS from the total sums of squares for the AR model . Next, the sums of squares for the OLS regression were obtained by subtracting the error sums of squares for the OLS from the total sums of squares for OLS.

To evaluate the effect of scale, the autoregressive models were calculated for 5 levels of aggregation. MSAVI values were averaged over 3, 5, 7 and 11 m^2 blocks using SURFER (Golden Software, Boulder CO, USA). A calibration was also carried out using Ordinary Least Squares Regression (OLS) where spatial relationships were not taken into account. In order to determine the spatial weighting matrix, D, we created individual neighbor weight matrices from first and second order Delaunay neighbors using the Spatial Statistics Toolbox (Pace and Barry, 2000). The nearest single neighbor was used.

3.0 RESULTS

3.0 SPATIAL AUTOREGRESSIVE MODEL

The spatial autregressive parameter for LAI, α, was essentially the same for all the levels of averaging, about 0.16 (Table 1) and it explains a significant proportion of the error ($p<0.01$). The range of spatial autocorrelation for the measured LAI data and reflectances was approximately 18-20 m (data not shown). This was within the range of the sampling distances for LAI. This was probably why the spatial correlation parameter, α, was small.

The relationship between LAI and MSAVI is not strong and the lowest root mean square error occurs for the 5 and 7

Table 1 Values of the Spatial Autoregression Coefficient, α, and Sums of Squares for the Various Components of Spatial Autogression of LAI on MSAVI.values

		Sums of Squares					
Block size (m2)	Alpha	Alpha	Regression	Error	Total	Mean square error	F_{alpha}
1	0.17	0.609	1.135	5.571	7.315	0.076	8.05
3	0.16	0.608	1.552	5.155	7.315	0.070	8.73
5	0.16	0.608	1.721	4.985	7.315	0.067	9.02
7	0.16	0.608	1.673	5.033	7.315	0.068	8.89
11	0.17	0.609	1.391	5.315	7.315	0.073	8.39

m block averages. The data in Figure 3 show that the high values of LAI were not represented well by the MSAVI

[**] Spatial Statistics Toolbox 1.1. R. Kelley Pace and R. Barry, available at: http:\\spatial-statistics.com

index of reflectances. This could be due to saturation of the reflectances used for MSAVI at high LAI. The weak relationship is also attributable to other factors. These include excessive pitching and yawing movement of the airplane due to windy conditions. The geographic positioning system located onboard the aircraft did not appear to account for all the movement of the plane. Furthermore, there were not a sufficient number of LAI values in the low range of LAI. Thus most of the measured values are clustered around the mean which makes it difficult to calibrate the full range of MSAVI values.

Table 2 Spatial Auto Regression (SAR) and Ordinary Least Squares Regression (OLS) Coefficients for the Linear Relationship Between MSAVI and LAI. *Spatial* Is the Coefficient for the Product of MSAVI and the Spatial Weight Matrix, D (B_2 in Eq. [1]).

	Spatial Autoregression				Ordinary Least Squares Regression		
Block size m2	Intercept	Slope	Spatial	r^2	Intercept	Slope	r^2
1	-0.315	-6.248	0.599	0.27	-0.401	-6.855	0.16
3	-0.870	-7.915	0.610	0.32	-0.999	-8.580	0.21
5	-1.160	-8.716	0.592	0.35	-1.298	-9.423	0.24
7	-1.152	-8.644	0.552	0.34	-1.272	-9.334	0.23
11	-0.926	-7.917	0.535	0.30	-1.020	-8.601	0.19

Compared to OLS regression alone for the 5-m block averages, the r^2 increased from 24 to 35% when spatial relationships were taken into account (Table 2). The correlation between the measured and estimated LAI is improved with SAR but the differences are not large. There appears to be a tighter clustering of the data about the 1:1 line for the SAR predicted values as compared to the OLS predicted (Figure 3).

The best correlation for the MSAVI-LAI relationship was at the scale of 5 to 7-m block averages. This was probably due to several reasons. First, the averaging of reflectance reduces noise. Second, the LAI instrument measures LAI in the crop to a distance that is about equal to the height of the crop (LICOR Co.) which was about 2 m at the time of measurement. This would result in a diameter of about 4 m around the crop that is being measured. This is similar to the 5-m block average. The magnitude of the both the slope and intercept of the relationship between MSAVI and LAI increased as block size increased up to the optimum block size of 7 m. The OLS slope and intercepts were also larger than the SAR estimates.

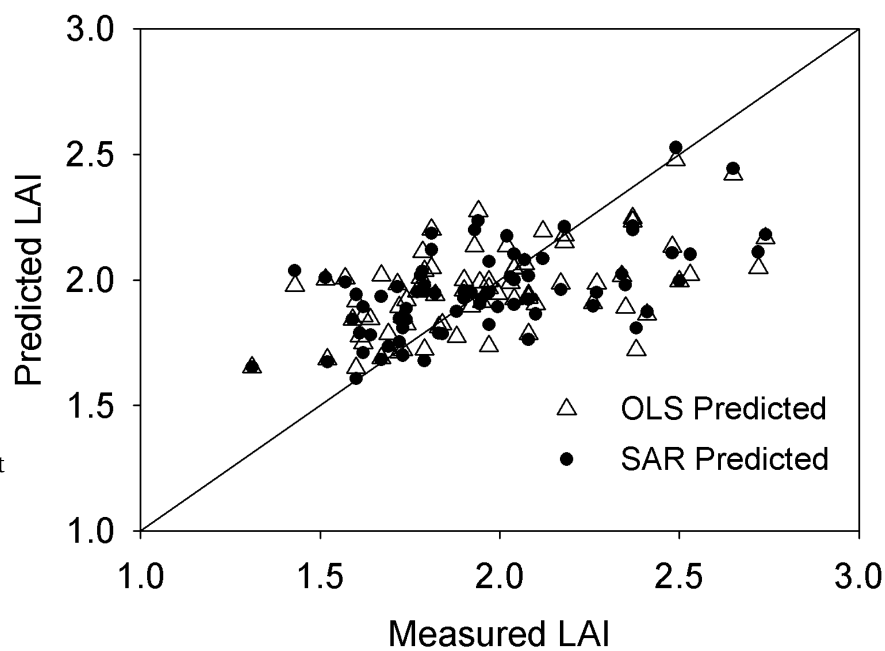

Figure A.3 Relationship between measured LAI and LAI predicted using a spatial autoregressive model and an ordinary least squares model. Data from 7 m^2 block averages are shown.

This is because the autoregressive parameter, *Spatial*, (Table 2) also accounts for some of the variance of MSAVI with LAI. The SAR parameter, *Spatial*, of 0.592 means that the neighboring value of MSAVI contributes 0.592 of it's value to predicting LAI while the MSAVI value at the measurement location contributes approximately -8 times it's value. The value *Spatial*, accounts for spatial autocorrelation in MSAVI while the value, α, accounts for spatial autocorrelation in LAI and spatial cross-correlation between MSAVI and LAI.

4.0 SUMMARY AND CONCLUSIONS

This study showed the importance of spatial information in calibration equations and how the spatial variance structure of the LAI and reflectance measurements can be used to improve estimates of LAI from reflectance data. The AR model predicted LAI at the measured locations with less error and bias than the OLS method did. The spatial correlations accounted for a significant portion of error in the autoregressive models. The spatial autocorrelation parameter, α, was small, about 0.16. This was probably because the sampling range of LAI was near its correlation length. It could be advantageous to sample LAI at more closely spaced intervals and thus increase the value of the autocorrelation parameter. This would allow the spatial autoregression to better correct predicted values of LAI by a non-parametric estimate of the error on nearby observations (Pace and Gilley, 1997). This would help offset errors in georectification of reflectances, and errors in measurement of LAI. Futhermore, the calibration would not provide a global fit of the data but rather a more local one. One possible disadvantage of this method would be the uncertainty of the consistency of the spatial autocorrelation over time.

5.0 REFERENCES

Atkinson PM and Curran PJ, "Choosing an appropriate spatial resolution for remote sensing investigations," *PE-and-RS,-Photogrammetric-Engineering-and-Remote-Sensing*, Vol. 63, No. 12, pp. 1345-1351, 1997.

Benedetti R and Rossini P, "On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna," *Remote-Sensing-of-Environment*, Vol. 45, No. 3, pp. 311-326, 1993.

Braam, B., J. Okkonen, M. Aikio, K. Makisara and J. Bolton, "Design and first test results of the Finnish airborne image spectrometer for different applications, AISA," In *Proc. Of Imaging Spectrometry of the Terrestrial Environment*, eds, G. Vane, Orlando, FL, SPIE Vol. 1937, pp. 142-151, April 14-15, 1993.

Chica-Olmo, M and F. Abarca-Hernandez. "Computing geostatistical image texture for remotely sensed data classification," *Computers in Geosciences* Vol. 26, No. 4, pp. 373-383, 2000.

Griffith, D.A, "A numerical simplification of estimating parameters of spatial autoregressive models." In: *Spatial Statistics: Past, Present, and Future,* ed., D.A. Griffith, Institute of Mathematical Geography, Ann Arbor, MI, pp.185-196, 1990.

Griffith, D.A. "Estimating spatial autoregressive model parameters with commercial statistical software," *Geographical Analysis*, Vol. 20, pp. 176-186. 1988.

Kirby, R., E. Matthews, and M. Bailey. *Soil survey of Prince Georges County, Maryland.* USDA-SCS. U.S. Govt. Printing Office, Washington, DC. 1967.

Long, D.S., "Spatial autoregression modeling of site-specific wheat," *Geoderma*. Vol. 85, pp. 181-197, 1998.

Pace, K.R. and R. Barry, "Quick computation of regressions with a spatially autoregressive dependent variable," *Geogr. Analysis*, Vol. 29, pp. 232-247, 1997.

Pace, R. K. and O.W. Gilley, "Using the spatial configuration of the data to improve estimation," *J. of Real Estate*

Finance and Economics, Vol. 14, No. 3, pp. 333-340, 1997.

Qi, J., Chehbouni, A., Huete, A. R., and Kerr, Y. H., "Modified Soil Adjusted Vegetation Index (MSAVI)," *Remote Sensing of Environment*, Vol. 48, pp. 119-126. 1994

6.0 ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Dr. Kelley Pace of Louisiana State Univ. for many helpful comments regarding spatial autocorrelation and for assistance with the Spatial Statistics Toolbox software.